

## WEST Search History





DATE: Tuesday, December 20, 2005

<u>Hide?</u>	<u>Set Name</u>	<u>Query</u>	<u>Hit Count</u>
	<i>DB=PGPB,USPT,EPAB; PLUR=YES; OP=ADJ</i>		
<input type="checkbox"/>	L40	L39 and L36	20
<input type="checkbox"/>	L39	L38 or L37	6033
<input type="checkbox"/>	L38	alpha chain	4386
<input type="checkbox"/>	L37	a-chain	1821
<input type="checkbox"/>	L36	L34 and L18	67
<input type="checkbox"/>	L35	L34 and L13	1506
<input type="checkbox"/>	L34	L32 or L31	16365
<input type="checkbox"/>	L33	L32or L31	0
<input type="checkbox"/>	L32	(530/395)![CCLS]	2138
<input type="checkbox"/>	L31	(514/12  514/2)![CCLS]	14501
<input type="checkbox"/>	L30	L29 and L13	4
<input type="checkbox"/>	L29	("5629327" "5639725" "5733876" "5792845" "5854205" "5981484").PN.	6
<input type="checkbox"/>	L28	L27 not @ay>2000	87
<input type="checkbox"/>	L27	L26 and L23	361
<input type="checkbox"/>	L26	L25 or L24	44209
<input type="checkbox"/>	L25	L12.clm.	18863
<input type="checkbox"/>	L24	L11.clm.	35604
<input type="checkbox"/>	L23	L22 and L12	559
<input type="checkbox"/>	L22	L19 and L11	651
<input type="checkbox"/>	L21	L13 and L11	11717
<input type="checkbox"/>	L20	L16 and L19	1
<input type="checkbox"/>	L19	L18 or L17	2132
<input type="checkbox"/>	L18	L13.ab.	943
<input type="checkbox"/>	L17	L13.clm.	1504
<input type="checkbox"/>	L16	L14 and L13	137
<input type="checkbox"/>	L15	L14 abd L13	0
<input type="checkbox"/>	L14	(liang or sim or dey).in.	11600
<input type="checkbox"/>	L13	(hepatocyte growth factor) or hgf or (scatter factor) or SF	35351
<input type="checkbox"/>	L12	angiogen\$ or prolifer\$ or metasta\$	134552
<input type="checkbox"/>	L11	cancer\$ or tumor\$ or neoplas\$	179243
<input type="checkbox"/>	L10	L9 not @py>1999	26

<input type="checkbox"/>	L9	L7 and L3	543
<input type="checkbox"/>	L8	L7 not @py>1999	26
<input type="checkbox"/>	L7	L6 and L4	543
<input type="checkbox"/>	L6	L3 with domain	1032
<input type="checkbox"/>	L5	L4 and L3	798
<input type="checkbox"/>	L4	angiogenesis	21045
<input type="checkbox"/>	L3	kringle	1439
<input type="checkbox"/>	L2	L1 and label\$	1
<input type="checkbox"/>	L1	5460950.pn.	1

END OF SEARCH HISTORY

OM protein - protein search, using sw model

Run on: November 18, 2005, 16:39:26 ; Search time 237 Seconds  
(without alignments)  
3450.143 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	9514	100.0	1861	3	AA90350	Aay90350 Drosophil
2	9482	99.7	1954	4	ABB62757	Abb62757 Drosophil
3	1360	14.3	3477	9	ADX83191	Adx83191 Human TEG
4	1360	14.3	3477	9	ADY19286	Ady19286 PRO polyp
5	1317	13.8	2000	9	ADY17682	Ady17682 PRO polyp
6	1278.5	13.4	1415	9	AEA19933	Aea19933 Novel hum
7	868	9.1	1142	9	ADY20048	Ady20048 PRO polyp
8	755	7.9	1277	8	ADQ65753	Adq65753 Novel hum

9	732	7.7	1752	8	ADT71531	Adt71531	Human	CGD
10	672.5	7.1	898	7	ADM04714	Adm04714	Human	pro
11	541	5.7	934	5	ABP64732	Abp64732	Human	pro
12	408.5	4.3	726	4	AAB92933	Aab92933	Human	pro
13	408.5	4.3	726	6	ABR48183	Abr48183	Human	bla
14	408.5	4.3	726	8	ADN04938	Adn04938	Antipsori	
15	319.5	3.4	526	4	AAB92906	Aab92906	Human	pro
16	303.5	3.2	2779	4	ABB62371	Abb62371	Drosophil	
17	302.5	3.2	4576	8	ADO55184	Ado55184	Protein #	
18	300.5	3.2	1855	6	ABR47539	Abr47539	Breast ca	
19	300.5	3.2	1855	8	ADN04056	Adn04056	Antipsori	
20	300	3.2	1374	8	ADU20432	Adu20432	A. thalia	
21	300	3.2	1381	8	ADU20579	Adu20579	A. thalia	
22	294	3.1	4684	8	ADH09904	Adh09904	Human	hos
23	294	3.1	4684	8	ADH09906	Adh09906	Human	hos
24	292	3.1	2000	9	AEA24046	Aea24046	Human	PRO
25	291	3.1	1742	7	ADB82767	Adb82767	Human	pro
26	290.5	3.1	285	4	AAB92905	Aab92905	Human	pro
27	290.5	3.1	285	8	ADQ18923	Adq18923	Human	sof
28	284.5	3.0	1805	6	ABU35589	Abu35589	Protein e	
29	284.5	3.0	1805	7	ABO23572	Abo23572	Mycoplasma	
30	282.5	3.0	1584	8	ABM80133	Abm80133	Tumour-as	
31	277	2.9	1818	6	ABU36321	Abu36321	Protein e	
32	277	2.9	1898	2	AAAY30795	Aay30795	A human t	
33	277	2.9	1898	7	ADD48869	Add48869	Human	Pro
34	275.5	2.9	2688	4	AAM40883	Aam40883	Human	pol
35	271.5	2.9	1975	4	ABB62094	Abb62094	Drosophil	
36	270	2.8	2383	5	ABG65631	Abg65631	Human	bre
37	270	2.8	2442	9	AEB13339	Aeb13339	c-NAP1 SE	
38	269.5	2.8	2663	4	AAM39097	Aam39097	Human	pol
39	269	2.8	1525	6	AAE32042	Aae32042	Human	ves
40	268	2.8	2482	2	AAR72826	Aar72826	Human	mit
41	268	2.8	2482	2	AAW23996	Aaw23996	Human	mit
42	268	2.8	3113	7	ADN95402	Adn95402	Human	BEC
43	268	2.8	3113	8	ADK70220	Adk70220	Human	oes
44	268	2.8	3113	8	ADL72172	Adl72172	Human	sol
45	268	2.8	3113	8	ADQ18045	Adq18045	Human	sof

# ALIGNMENTS

## RESULT 1

AAAY90350

ID AAY90350 standard; protein; 1861 AA.

XX

AC AAY90350;

XX

DT 04-DEC-2000 (first entry)

XX

DE Drosophila Asp protein sequence.

XX

KW Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor; tumour cell.

XX

OS Drosophila sp.

XX

```

PN      WO200052478-A1.
XX
PD      08-SEP-2000.
XX
PF      03-MAR-2000; 2000WO-GB000785.
XX
PR      04-MAR-1999; 99GB-00005007.
XX
PA      (UYDU-) UNIV DUNDEE.
XX
PI      Glover DM, Avides MDC;
XX
DR      WPI; 2000-594203/56.
DR      N-PSDB; AAA37761.
XX
PT      Use of Drosophila Asp polypeptide for identifying substances capable of
PT      disrupting microtubule organizing center integrity and use of the
PT      identified substances for inhibiting mitosis in tumor cell.
XX
PS      Claim 4; Page 43-44; 51pp; English.
XX
CC      This sequence represents the Drosophila Asp protein. The invention
CC      relates to the use of Drosophila Asp polypeptide (or its homologue, or
CC      fragment) capable of stimulating formation and/or maintenance of
CC      microtubule organising centres (MTOCs), in an assay for identifying a
CC      substance capable of disrupting MTOC integrity. Asp polypeptide or its
CC      homolog is useful for identifying a substance capable of disrupting MTOC
CC      integrity. Substances identified by the method can be used to inhibit
CC      mitosis, e.g. in tumour cells
XX
SQ      Sequence 1861 AA;

Query Match          100.0%; Score 9514; DB 3; Length 1861;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSPT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSPT 60

Qy      61 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTWRVTAPSRPAAWAHPPPQAPLVEKNVYKT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTWRVTAPSRPAAWAHPPPQAPLVEKNVYKT 120

Qy      121 PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRKGQATIFPDNLAAWPT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRKGQATIFPDNLAAWPT 180

Qy      181 PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT 240

Qy      241 PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD 300

Qy      301 LSLSPQTKYAIQGSMPNLEMKIRSIEQNRYEQEQIQIKAKDLNSSSSSEASLAGQQE 360

```

Db	301	LSLLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	360
Qy	361	FLFNHSEILAQSSRFNLHEVGRKSVKGS PVKNPHKRRSHELFS DAPS NESLYRNETVAI	420
Db	361	FLFNHSEILAQSSRFNLHEVGRKSVKGS PVKNPHKRRSHELFS DAPS NESLYRNETVAI	420
Qy	421	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKF KLAQTMSLMKKPATPRKVRD	480
Db	421	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKF KLAQTMSLMKKPATPRKVRD	480
Qy	481	TSIQPSVKLYDSELYMQTCINPD PFAATTTIDPFLASTMYLDEQAVDRHQADF KKWLNAL	540
Db	481	TSIQPSVKLYDSELYMQTCINPD PFAATTTIDPFLASTMYLDEQAVDRHQADF KKWLNAL	540
Qy	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLT KYRLET LRKA AVELFF	600
Db	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLT KYRLET LRKA AVELFF	600
Qy	601	SEQMRLPCSKVAVYVNKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLR LGLEVV FGE	660
Db	601	SEQMRLPCSKVAVYVNKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLR LGLEVV FGE	660
Qy	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHS LQKILFLLPFL	720
Db	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHS LQKILFLLPFL	720
Qy	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRR LGYVLQHRQTF	780
Db	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRR LGYVLQHRQTF	780
Qy	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVP AISRLQRI FNVKLALGALG	840
Db	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVP AISRLQRI FNVKLALGALG	840
Qy	841	EANFQLGGDIAAQDIVDGHREKTL SLLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	900
Db	841	EANFQLGGDIAAQDIVDGHREKTL SLLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	900
Qy	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFK TERTQA AII LQKFTRRYLAQKQLYQ	960
Db	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFK TERTQA AII LQKFTRRYLAQKQLYQ	960
Qy	961	SYHSIITIQRWWRAQQ LGRQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAETARLQRS	1020
Db	961	SYHSIITIQRWWRAQQ LGRQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAETARLQRS	1020
Qy	1021	QKQQAAS YIQMWRTYQLGRIQRHEFLRQRD LIMFVQRRMRSKWSMLEQRKEFQQLKRA	1080
Db	1021	QKQQAAS YIQMWRTYQLGRIQRHEFLRQRD LIMFVQRRMRSKWSMLEQRKEFQQLKRA	1080
Qy	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1140
Db	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1140
Qy	1141	RLRAIMKMREQRENYLR LRNASILVQKRYRMQQMIQDRNAYLRTRKCI INVQRRWRATL	1200

Db	1141	RLRAIMKMRQRENYLRLRNASILVQKRYMRQMIQDRNAYLRTKCIINVQRRWRATL	1200
Qy	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1260
Db	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1260
Qy	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTAQAVSCLQMHWRNHLLRKRENSFLQL	1320
Db	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTAQAVSCLQMHWRNHLLRKRENSFLQL	1320
Qy	1321	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQVVLVYQKQREAIK	1380
Db	1321	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQVVLVYQKQREAIK	1380
Qy	1381	VQRRYRGNLEMRKQIEVYQKQKQAVIRLQKWWRSIRDMRLCKAGYRRIRLSSLSIQRKWR	1440
Db	1381	VQRRYRGNLEMRKQIEVYQKQKQAVIRLQKWWRSIRDMRLCKAGYRRIRLSSLSIQRKWR	1440
Qy	1441	ATVQARRQREIFLSTIRKVRIMQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML	1500
Db	1441	ATVQARRQREIFLSTIRKVRIMQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML	1500
Qy	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMIEQRNCF	1560
Db	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMIEQRNCF	1560
Qy	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1620
Db	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1620
Qy	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLL	1680
Db	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLL	1680
Qy	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1740
Db	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1740
Qy	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ	1800
Db	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ	1800
Qy	1801	NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDEFGIIRYSPYTFISSVYAFDTILCKLQIDM	1860
Db	1801	NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDEFGIIRYSPYTFISSVYAFDTILCKLQIDM	1860
Qy	1861	F 1861	
Db	1861	F 1861	

# RESULT 2

ABB62757

ID ABB62757 standard; protein; 1954 AA.

XX

AC ABB62757;

XX

DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 15063.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL06860.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 15063; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1954 AA;

Query Match 99.7%; Score 9482; DB 4; Length 1954;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1855; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MELVWSPVLEVACKETLQLIDNRNFRKEVMILKSKSNQPVKNPRKFPTVGKTLQLKSPT 60  
 |||  
 Db 94 MELVWSPVLEVACKETLQLIDNRNFRKEVMILKSKSNQPVKNPRKFPTVGKTLQLKSPT 153  
 Qy 61 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRVVTAPSRPAAWAHPPPQAPLVEKNVYKT 120  
 |||  
 Db 154 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRVVTAPSRPAAWAHPPPQAPLVEKNVYKT 213  
 Qy 121 PQEFPVYISPQPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT 180



Db	214	 PQEEPVIISPQPRSLKENLSPMTPGNLLDVIDNLRFPTLTETRGKGQATIFPDNLAAWPT	273
Qy	181	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT	240
Db	274	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT	333
Qy	241	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD	300
Db	334	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD	393
Qy	301	LSSLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	360
Db	394	LSSLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	453
Qy	361	FLFNHSEILAQSSRFLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESLYRNETVAI	420
Db	454	FLFNHSEILAQSSRFLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESLYRNETVAI	513
Qy	421	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	480
Db	514	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	573
Qy	481	TSIQPSVKLYDSELYMQTCINPDPPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL	540
Db	574	TSIQPSVKLYDSELYMQTCINPDPPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL	633
Qy	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF	600
Db	634	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF	693
Qy	601	SEQMRLPCSKVAVYVNKQALRIRSDRNLHLDVVMQRTILELLLCFNPLWLRLGLEVVFE	660
Db	694	SEQMRLPCSKVAVYVNKQALRIRSDRNLHLDVVMQRTILELLLCFNPLWLRLGLEVVFE	753
Qy	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSLOKILFLLPFL	720
Db	754	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSLOKILFLLPFL	813
Qy	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSELLANIGDITRELRLGYVLQHRQTF	780
Db	814	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSELLANIGDITRELRLGYVLQHRQTF	873
Qy	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPASRLQRI FNVKLALGALG	840
Db	874	LDEFDYAFNNLAVDLRDGVRLTRVMEVILLRDDLTRQLRVPASRLQRI FNVKLALGALG	933
Qy	841	EANFQLGGDIAAQDIVDGHREKTLSSLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	900
Db	934	EANFQLGGDIAAQDIVDGHREKTLSSLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	993
Qy	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAIILQKFTRRYLAQKQLYQ	960
Db	994	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAIILQKFTRRYLAQKQLYQ	1053
Qy	961	SYHSIITIQRWWRAQQQLGRQHRQRFVELREAAIFLQRIWRRRLFAKKLLAAAETARLQRS	1020

Db	1054	SYHSIITIQRWWRAQQQLGRQHRQRFVELREAAIFLQRIWRRRLF	FAKKLLAAAE	TARLQRS	1113
Qy	1021	QKQQAAASYIQMQWRTYQLGRIQRHEFLRQRD	LIMFVQRRMRSKWSM	LEQRKEFQQLKRA	1080
			:		:
Db	1114	QKQQAAASYIQMQWRSYQLGRIQRQQFLRQRD	LIMFVQRRMRSKWSM	LEQRKEFQQLKRA	1173
Qy	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI	QMRIDRNHYYS	LRKNVICLQQ	1140
Db	1174	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI	QMRIDRNHYYS	LRKNVICLQQ	1233
Qy	1141	RLRAIMKMREQRENYLRNRNASILVQKRYRMRQQMIQDRNAYLR	TRKCIINVQRRWRATL		1200
Db	1234	RLRAIMKMREQRENYLRNRNASILVQKRYRMRQQMIQDRNAYLR	TRKCIINVQRRWRATL		1293
Qy	1201	QMRERKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVT	LVVQKRRRALLQMRKER		1260
Db	1294	QMRERKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVT	LVVQKRRRALLQMRKER		1353
Qy	1261	QEYLHLREVTIKLQRRFHAQKSMRFRMAKYRGTOAAVSC	LQMHWRNHLLRKRENS	SFLQL	1320
Db	1354	QEYLHLREVTIKLQRRFHAQKSMRFRMAKYRGTOAAVSC	LQMHWRNHLLRKRENS	SFLQL	1413
Qy	1321	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQV	VLYQKQREAIK		1380
Db	1414	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQV	VLYQKQREAIK		1473
Qy	1381	VQRRYRGNLEMRKQIEVYQKQKQAVIRLQKWWRSIRDMLCKAGYRRIRL	SSLSIQRKWR		1440
Db	1474	VQRRYRGNLEMRKQIEVYQKQKQAVIRLQKWWRSIRDMLCKAGYRRIRL	SSLSIQRKWR		1533
Qy	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARC	AML		1500
Db	1534	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARC	AML		1593
Qy	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMIEQRNCF			1560
Db	1594	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQKFRGKRLMIEQRNCF			1653
Qy	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ			1620
Db	1654	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ			1713
Qy	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVP	HLL		1680
Db	1714	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSRLDRLSRTVP	HLL		1773
Qy	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM			1740
Db	1774	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM			1833
Qy	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ			1800
Db	1834	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ			1893
Qy	1801	NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM			1860
Db	1894	NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM			1953

Qy 1861 F 1861  
|  
Db 1954 F 1954

RESULT 3

ADX83191

ID ADX83191 standard; protein; 3477 AA.

XX

AC ADX83191;

XX

DT 05-MAY-2005 (first entry)

XX

DE Human TEG14 polypeptide SEQ ID NO 73.

XX

KW cytostatic; protein purification; diagnosis; cancer; cytostatic;  
KW neoplasm; respiratory disease; lung tumor; gastrointestinal disease;  
KW stomach tumor; colon tumor; hepatic tumor; selectable marker; TEG.

XX

OS Homo sapiens.

XX

PN WO2005014818-A1.

XX

PD 17-FEB-2005.

XX

PF 06-AUG-2004; 2004WO-JP011650.

XX

PR 08-AUG-2003; 2003JP-00290704.

XX

PA (PERS-) PERSEUS PROTEOMICS INC.

PA (CHUS ) CHUGAI SEIYAKU KK.

PA (ABUR/) ABURATANI H.

XX

PI Aburatani H, Hippo Y, Taniguchi H, Chen YX, Ishikawa S;

PI Fukumoto S, Shimamura T, Kamimura N, Guo YQ, Yamamoto S, Ito Y;

PI Ito H, Ohtomo T;

XX

DR WPI; 2005-173106/18.

DR N-PSDB; ADX83135.

XX

PT Novel protein encoded by any one of TEG1 to TEG64, useful for diagnosing  
PT and treating cancer e.g. lung, hepatic, stomach, colon or pancreatic  
PT cancer.

XX

PS Disclosure; SEQ ID NO 73; 462pp; Japanese.

XX

CC The invention describes a protein (I) encoded by a gene having a  
CC nucleotide sequence chosen from any one of 65 fully defined 418-19341  
CC base pair sequences (SEQ ID No. 1-65) (TEG 1-64) or their fragments. Also  
CC described are: a protein (II) encoded by a gene having a nucleotide  
CC sequence chosen from SEQ ID No. 1, 2, 28, 29, 30, 31, 32, 51, 52, 60 and  
CC 61 or their fragments; a protein (III) encoded by a gene having a  
CC nucleotide sequence chosen from any one of SEQ ID No. 3-13, 22-27 and 33-  
CC 55 or their fragments; protein (IV) encoded by a gene having a nucleotide  
CC sequence chosen from SEQ ID No. 3, 7, 20, 21 and 46-50 or their fragments  
CC ; a protein (V) encoded by a gene having a nucleotide sequence chosen

CC from SEQ ID No. 14-19, 43-45, 56-59 and 62-65 or their fragments; an  
 CC antibody (VI) that specifically recognizes any one of (I)-(V) or their  
 CC fragments; a polynucleotide (VII) complementary to the nucleotide  
 CC sequence of any one of SEQ ID No. 1-65 or a polynucleotide sequence  
 CC capable of hybridizing with SEQ ID No. 1-65; a polynucleotide (VIII)  
 CC comprising at least 12 consecutive nucleotides in any one SEQ ID No. 1-65  
 CC or a polynucleotide sequence capable of hybridizing with this nucleotide;  
 CC a composition (C1) for diagnosing or treating lung cancer; a composition  
 CC (C2) for diagnosing or treating stomach cancer; composition (C3) for  
 CC diagnosing or treating colon cancer; a composition (C4) for diagnosing or  
 CC treating hepatic cancer; a vector (IX) comprising (VII) or (VIII); a cell  
 CC (X) comprising (IX); identifying (M1) a compound having anticancer  
 CC activity; and diagnosing (M2) cancer. Proteins (I)-(V), an antibody (VI)  
 CC that specifically binds the proteins and polynucleotides (VII) and (VIII)  
 CC are useful for diagnosing cancer. (M2) comprising measuring the  
 CC expression level of (I)-(V) or (VII) or (VIII), or obtaining sample  
 CC (blood serum or plasma), and detecting C20orf102 protein in the obtained  
 CC sample is also useful for diagnosing cancer such as lung cancer, hepatic  
 CC cancer or pancreatic cancer, where the C20orf102 protein is a secreted or  
 CC extracellular C20orf102 protein, which is detected using an antibody  
 CC which recognizes C20orf102 protein. A composition (C1) comprising protein  
 CC (II) is useful for diagnosing or treating lung cancer. A composition (C2)  
 CC comprising protein (III) is useful for diagnosing or treating stomach  
 CC cancer. A composition (C3) comprising protein (IV) is useful for  
 CC diagnosing or treating colon cancer. A composition (C4) comprising (V) is  
 CC useful for diagnosing or treating hepatic cancer. (M1) comprising  
 CC contacting a cultured human cell with a test compound and identifying a  
 CC compound that causes a change in expression level of the gene which  
 CC contains the nucleotide sequence in any one of SEQ ID No. 1-65 is useful  
 CC for identifying a compound having anticancer activity. A vector (IX) or  
 CC cell (X) is useful for producing (I) or (VI). An antibody (VI) is useful  
 CC as a cancer diagnostic marker. This is the amino acid sequence of a human  
 CC TEG polypeptide.

XX

SQ Sequence 3477 AA;

Query Match 14.3%; Score 1360; DB 9; Length 3477;  
 Best Local Similarity 22.7%; Pred. No. 4.6e-93;  
 Matches 480; Conservative 360; Mismatches 630; Indels 640; Gaps 68;

Qy	5	WSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSPTGAGK	64
		::  :  : :  :    ::  :   : :	
Db	110	WTPLKEGRVREIMTFLVN-DVLKHQAILLGNAEEQKKKKRSLWDTI-----	154
Qy	65	TMKSVVSAAVQQKKRMS-----AAAAPPSKQTWRVTAPSRPAAWAHPPPQAPLVEKNVYK	119
		:  : :  :  ::    :  :	
Db	155	-KKKKISASTSHNRRVSNIQNVNKTFSVSQKVDRVRSPLQACENLAMNEGGPPTENNSL-	212
Qy	120	TPQEEPVIYISP-----QPRSLK-----ENLSPMTPGNLLDV----IDNLRFT	157
		:  :         :: :    :  : :	
Db	213	ILEENKIPISPISPAFNECHGATCLPLSVRRSTTYSSSLHASENRELLNVHSANVSKVSFN	272
Qy	158	--PLTET-----RGKGQATIFPDNLAAWPTPTLKGNVKSCAN--DMRPRRITPDD	203
		:      : :        : :	
Db	273	EKAVTETSFNSVNVNGQRGENSKL-----SLTPNCSSTLNITQSQIHFLSPDS	320
Qy	204	LEDQPATNKTFDVKHSETINISLDTLDCSRID---GQPHTPLNKTTTIVHATHTRALACI	260

Db	321	F-----VNNSHGANNELELVTCCLSSDMFMKDNSQPVHLESTIAHEIYQKIL---	366
Qy	261	HEEEGPSPPRTPTKSAIHD---LKRDIKLVG-SPLRKYSESMKDLSLLSPQTKYAIQGSM	316
Db	367	-----SP-----DSFIKDNYGLNQDLESESVNPILSPNQFLKDNMAYMCTSQQTCKVPL	415
Qy	317	PNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQEFLFNHSEILAQSS-RF	375
Db	416	SNENSQVPQSPED---WRKSEVSPRIPECQGSKSPKAI FEELVEMKSNYYSEFIKQNNPKF	472
Qy	376	N-LHEVGRKSVKGSVPKNP-----HKRRS-----HELSEF	403
Db	473	SAVQDISSSHSHNKQPKRRPILSATVTKRKATCTRENQTEINKPKAKRCLNSAVGEHEKVI	532
Qy	404	SDAPSNESESL-----YRNE-----TVAISPPKKQ--RVEDTTLPRSAAP	439
Db	533	NNQKEKEDFHSYLPIDPILSKSKSYKNEVTPSSTTASVARKRSDGSMEDANV-RVAIT	591
Qy	440	ANASARS-SSAHAWPHAQ-----SKKFKLAQTMSLMKKP-----ATPRKV	478
Db	592	EHTEVREIKRIHFSPSEPKTSAVKKTKNVTTPISKRISNREKLNKKKTDLISIFRTPISK	651
Qy	479	RDTSIQPSVKLYDSELYMQTCINPDPAATTTID----PFLASTMYLDEQAVDRHQADFK	534
Db	652	TNKRTPKPIIAVAQSSL---TFIKP----LKTDIPRHPMPFAAKNMFYDERWKEKQEQGFT	704
Qy	535	KWLNALVSI PADLDADLN-NKIDVGKLFNEVRNKELVV---APTKEEQSMN-YLTKYRLE	589
Db	705	WWLNFILT-PDDFTVKTNI SEVNAATLLLGIEHQHKISVPRAPTKEEMSLRAYTARCLN	763
Qy	590	TLRKA AVELFFSEQMRLPCSKVAVYVNKQALRIRSDRNLHLDVVMQRTILELLLCFNPLW	649
Db	764	RLRRAACRLFTSEKMVKAIKKLEIEI EARRLIVRKDRHLWKDVGERQKVLNWLLSYNPLW	823
Qy	650	LRLGLEVVFG EKIQMQSNRDI VGLSTFILNRLFRN---KCEEQRYSKAYTLTEEYAETIK	706
Db	824	LRIGLETTYGELISLEDNSDVTGLAMFILNRLWNPDIAAEYRHP TVPHLYRDGHEEALS	883
Qy	707	KHSLQKILFLLPFLDQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRE	766
Db	884	KFTLKKLLLLVCFLDYAKISRLIDHDPCLFCKDAEFKASKEILLAFSRDFLSGEGDLSRH	943
Qy	767	LRRLG YVLQHRQTF LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPAISRL	826
Db	944	LGLLGLPVNHVQTPFDEFDAVTNLAVDLQCGVRLVRTMELLTQNWDL SKKLRI PAISRL	1003
Qy	827	QRIFNVKLALGALGEANFQL----GGDIAAQDIVDGHREKTL SLLWQLIYKFR-----	875
Db	1004	QKMHNVDIVLQVLKSRGIELSDEHGNTILSKDIVDRHREKTLRLLWKIAFAFQVDISLNL	1063
Qy	876	-----	875
Db	1064	DQLKEEIAFLKHTKSIKKTISLLSCHSDDLINKKKKGKRD SGSEFQYSENIKLMDWVNAV	1123
Qy	876	-----	875

Db 1124 CAFYNKKVENFTVSFSDGRVLCYLIHHYHPCYVPFDAICQRTTQTVECTQTGSVVLNSSS 1183  
 Qy 876 -----SPKFH----- 880  
 ||  
 Db 1184 ESDDSSLDMSLKAFDHENTSELYKELLENEKKNFHLVRSASVRDLGGIPAMINHSDMSNTI 1243  
 Qy 881 -----AAATVLQKWRRHWLHVVIQRRIRHKELMRRHRAAT 916  
 || ::| ||:: | :: ||:| | :||  
 Db 1244 PDEKVVITYLSFLCARLLDLRKEIRAARLIQTTWRKYKLKTDLK---RHQE---REKAAR 1297  
 Qy 917 VIQAVFRGHQMRKYVKLFKTERTQAAIILQKFTRRYLAQKQLYQ----- 960  
 :|| :: :: :| ||::||: || |||::|  
 Db 1298 IIQLAVINFLAKQRLR----KRVNAALVIQKYWRRVLAQRKLLMLKKEKLEKVQNKAAASL 1353  
 Qy 961 -----SYHSII-----TIQRWWRAQQLGROH 981  
 |:||| |||| ||| :|  
 Db 1354 IQGYWRRYSTRQRFLLKLYSIIQSRIIRMIIVTSYKRYLWATVTIQRHWRAYLRRKQD 1413  
 Qy 982 RQRFVELREAAIFLQ---RIWRRRLFAKKLLAAETAR-----LQRSQKQQAASIIQM 1032  
 :||: | : : : | | |::| :: | | :| :| :| :| ||  
 Db 1414 QQRYEMLKSSTLIQSMFRKWKQKMQSQVKATVILQRAFREWHLRKQAKEENSAIIQS 1473  
 Qy 1033 QWRTYQLGRIQRHEFLRQDLIMFVQRRMRSKWSMLEQRKEFQQLKRAAINIQQRWRAKL 1092  
 :| :: : ::| | :: :|:| | : :| ::| | : : ||: ::| |  
 Db 1474 WYRMHK----ELRKYIYIRSCVVIQKRFR----CFQAQKLYKRRKESILTIQKYKAYL 1525  
 Qy 1093 SMRKCNDYLLALRSSVLKVQA-----YRKATI-----QMRIDRNHYYS LRKN 1134  
 : :|| |:: ::|| | | : :|| || : :|:|  
 Db 1526 KGKIERTNYLQKRAAAIQLQAAFRRLKAHNLCRQIRAACVQSYWRRMQDRVFLNLKKT 1585  
 Qy 1135 VICLQQRLRAIMKMQRENYLRLRNASILVQKRYRMQQMIQDRNAYLRTRKCIINVQR 1194  
 :| : :| :: :||: | ::| |:::| :| : : :| :|| :| :|  
 Db 1586 II----KFQAHVRKHQQRQKYKKMKKAAVIIQTHFRAYIFAMKVLASYQKTRSAVIVLQS 1641  
 Qy 1195 RWRATLQMRERKNYLHLQT'TTKRIQIKFRA---KRE-----MKKQRAE 1235  
 :| : || |::| : :|| :|| |:| ||: | :  
 Db 1642 AYRG----MQARKMYIHILTSVIKIQSYRAYVSKKEFLSLKNATIKLQSTVKKMQTRKQ 1697  
 Qy 1236 FLQLKKVTLVVQKRRRALLQMRKERQEYLHLREVTIKLQ---RRFHAQKSMRFMRACYRG 1292  
 :| | : | :| : : :|:|: :|| |||| | : :| || |  
 Db 1698 YLHLRAAALFIQQCYRSKKIAAQKREEYMQMRESCIKLQAFVRGYLVKQMRQLR----- 1752  
 Qy 1293 TQAAVSCLQMHWRNHLRKRERNSFLQLRQAATLQRRYRARLNMILKLSYAQLKQAAI 1352  
 || || ::| :: | :|:: :| :| | | : | |:: |::||  
 Db 1753 --KAVISLQSYFR----MRKARQYYLKYKAIIVIQNYHAYKAQVNQRKNFLQVKKAAAT 1806  
 Qy 1353 TIQTRYRAKKAMQKQVVLVYQKQREAIKQRRYRGNLEMRKQIEVYQKQRAVIRLQKWW 1412  
 :| || | | | :|| | :|:| :|| |::| || |::|:|:|:|  
 Db 1807 CLQAAYRGYKVRQ----LIKQQSIAALKIQSAFRG---YNKRVK-YQSVLQSIKIQRWY 1858  
 Qy 1413 RSIRDMRLCKAGYRRIRLSSLSIQ---RKWRATVQARRQREIFLSTIRKVRMLQAFIRAT 1469  
 | : : : : : : :|:| | | : | ||: :  
 Db 1859 RAYKTLHDTRTHFLKTKAAVISLQSAYRGWKVRKQIRREHQ----- 1899  
 Qy 1470 LLMRQQRREFEMKRRAAVVIQRRFRARCAMLKARQDYQLIQSSVILVQRKFRANRSMKQA 1529  
 ||: || || | ||:: ::| :: :::|: || : :  
 Db 1900 -----AALKIQSAFR----MAKAQKQFRLFKTAALVIQQNFRAWTAGRKQ 1940

Qy	1530	RQEFVQLRTIAVHLQQKFRGK---RLMIEQRNCFQLLRCSMPGFQARARGFMARKRFQAL	1586
		:::   :    ::     :     ::   :   : : ::: :	
Db	1941	CMEYIELRHAVLVLSMWKGKTLRRQLRQHKCAIII-----QSYRMHVQQKKWKIM	1993
Qy	1587	MTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQGL-----LDIRKRIAQ	1631
		:   :  : :              : :	
Db	1994	-----KKAALLIQKYYRAYSIGREQNHLYLKTAAVVTLSAYRGMKVRKRIKD	2042
Qy	1632	LRQEAKAVNS	1641
		:   :	
Db	2043	CNKAAVTIQS	2052

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2005, 17:09:58 ; Search time 220 Seconds  
(without alignments)  
3534.456 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	9482	99.7	1954	6	US-11-097-143-15063	Sequence 15063, A
2	672.5	7.1	898	4	US-10-108-260A-3399	Sequence 3399, Ap
3	408.5	4.3	726	4	US-10-188-832-82	Sequence 82, Appl
4	348	3.7	1022	4	US-10-437-963-156007	Sequence 156007,
5	324	3.4	1407	5	US-10-732-923-3359	Sequence 3359, Ap
6	323.5	3.4	1549	5	US-10-732-923-3358	Sequence 3358, Ap
7	314	3.3	1909	5	US-10-732-923-3341	Sequence 3341, Ap
8	303.5	3.2	2779	6	US-11-097-143-13905	Sequence 13905, A
9	300.5	3.2	1855	4	US-10-177-293-315	Sequence 315, App
10	291	3.1	1742	4	US-10-012-697-1548	Sequence 1548, Ap



11	291	3.1	1742	5	US-10-779-543-23548	Sequence 23548, A
12	290.5	3.1	285	5	US-10-723-860-1742	Sequence 1742, Ap
13	284.5	3.0	1805	3	US-09-820-843A-73	Sequence 73, Appl
14	284.5	3.0	1805	4	US-10-282-122A-63513	Sequence 63513, A
15	284.5	3.0	1805	5	US-10-732-923-3303	Sequence 3303, Ap
16	277	2.9	1818	4	US-10-282-122A-64245	Sequence 64245, A
17	277	2.9	1818	5	US-10-732-923-3306	Sequence 3306, Ap
18	271.5	2.9	1975	6	US-11-097-143-13074	Sequence 13074, A
19	270	2.8	2383	4	US-10-082-830-260	Sequence 260, App
20	269	2.8	1525	5	US-10-475-476-8	Sequence 8, Appli
21	268	2.8	3113	5	US-10-723-860-862	Sequence 862, App
22	268	2.8	3113	5	US-10-756-149-4927	Sequence 4927, Ap
23	267.5	2.8	2663	5	US-10-723-860-749	Sequence 749, App
24	266	2.8	3114	5	US-10-751-736-99	Sequence 99, Appl
25	265.5	2.8	2633	5	US-10-450-763-36864	Sequence 36864, A
26	262.5	2.8	2568	5	US-10-828-985A-7	Sequence 7, Appli
27	260	2.7	1979	4	US-10-205-823-419	Sequence 419, App
28	260	2.7	1979	6	US-11-051-454-419	Sequence 419, App
29	257	2.7	2543	5	US-10-828-985A-9	Sequence 9, Appli
30	255.5	2.7	2017	5	US-10-450-763-36660	Sequence 36660, A
31	254.5	2.7	1837	4	US-10-369-493-22734	Sequence 22734, A
32	252.5	2.7	2115	5	US-10-977-955-33	Sequence 33, Appl
33	251.5	2.6	1792	6	US-11-097-143-40116	Sequence 40116, A
34	250.5	2.6	1151	4	US-10-408-765A-994	Sequence 994, App
35	250.5	2.6	1416	4	US-10-295-027-446	Sequence 446, App
36	250	2.6	2101	5	US-10-977-955-32	Sequence 32, Appl
37	250	2.6	2101	5	US-10-723-681-18	Sequence 18, Appl
38	249.5	2.6	1439	5	US-10-754-079-157	Sequence 157, App
39	249.5	2.6	2503	5	US-10-828-985A-11	Sequence 11, Appl
40	247.5	2.6	2246	5	US-10-450-763-36209	Sequence 36209, A
41	244.5	2.6	1401	4	US-10-408-765A-2125	Sequence 2125, Ap
42	239.5	2.5	4365	5	US-10-472-928-3660	Sequence 3660, Ap
43	238.5	2.5	1489	6	US-11-097-143-6636	Sequence 6636, Ap
44	238.5	2.5	1893	4	US-10-408-765A-1696	Sequence 1696, Ap
45	238.5	2.5	3674	4	US-10-291-265-454	Sequence 454, App

#### ALIGNMENTS

##### RESULT 1

US-11-097-143-15063

; Sequence 15063, Application US/11097143

; Publication No. US20050208558A1

; GENERAL INFORMATION:

; APPLICANT: Venter, J. Craig

; APPLICANT: et al.

; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

; TITLE OF INVENTION: DROSOPHILA GENES.

; FILE REFERENCE: CL000728

; CURRENT APPLICATION NUMBER: US/11/097,143

; CURRENT FILING DATE: 2005-04-04

; PRIOR APPLICATION NUMBER: 60/157,832

; PRIOR FILING DATE: 1999-10-05

; PRIOR APPLICATION NUMBER: 60/160,191

```
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15063
;   LENGTH: 1954
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-15063
```

Qy	1	MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVVGKTLQLKSPT	60
Db	94	MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVVGKTLQLKSPT	153
Qy	61	GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTWRVTAPSRPAAWAHPPPQAPLVEKNVYKT	120
Db	154	GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTWRVTAPSRPAAWAHPPPQAPLVEKNVYKT	213
Qy	121	PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT	180
Db	214	PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT	273
Qy	181	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT	240
Db	274	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPHT	333
Qy	241	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD	300
Db	334	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSESMKD	393
Qy	301	LSLLSPQTKYAIQGSMPNPNLNMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	360
Db	394	LSLLSPQTKYAIQGSMPNPNLNMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	453
Qy	361	FLFNHSEILAQSSRFNLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI	420
Db	454	FLFNHSEILAQSSRFNLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI	513
Qy	421	SPPKKQRVEDTTLPESAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	480

Db	514	SPPKKQRVEDTTLP	SAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	573
Qy	481	TSIQPSVKLYDSELYMQTCINPDPFAATTTIDPFLASTMYLDEQAVDRHQADF	FKKWLNAL	540
Db	574	TSIQPSVKLYDSELYMQTCINPDPFAATTTIDPFLASTMYLDEQAVDRHQADF	FKKWLNAL	633
Qy	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLT	KYRLET	600
Db	634	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLT	KYRLET	693
Qy	601	SEQMRLPCSKVAVYVNKQALRIRSDRNLHLDVVMQRTILELLLCFNPLWLR	LGLEVVFE	660
Db	694	SEQMRLPCSKVAVYVNKQALRIRSDRNLHLDVVMQRTILELLLCFNPLWLR	LGLEVVFE	753
Qy	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHS	LQKILFLLPFL	720
Db	754	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHS	LQKILFLLPFL	813
Qy	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITREL	RRLGYVLQHRQTF	780
Db	814	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITREL	RRLGYVLQHRQTF	873
Qy	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVP	PAISRLQRI	840
Db	874	LDEFDYAFNNLAVDLRDGVRLTRVMEVILLRDDLTRQLRVP	PAISRLQRI	933
Qy	841	EANFQLGGDIAAQDIVDGHREKTL	SLLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	900
Db	934	EANFQLGGDIAAQDIVDGHREKTL	SLLWQLIYKFRSPKFHAAATVLQKWWRRHWLHVVIQ	993
Qy	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAI	ILQKFTRRYLAQKQLYQ	960
Db	994	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAI	ILQKFTRRYLAQKQLYQ	1053
Qy	961	SYHSIITIQRWWRAQQLGRQHRQRFVELREAAIFLQRIWRRRL	FAKKLLAAAETARLQRS	1020
Db	1054	SYHSIITIQRWWRAQQLGRQHRQRFVELREAAIFLQRIWRRRL	FAKKLLAAAETARLQRS	1113
Qy	1021	QKQQAAS	YIQMQWRTYQLGRIQRHEFLRQRD	1080
Db	1114	QKQQAAS	YIQMQWRSYQLGRIQRQFLRQRD	1173
Qy	1081	AINIQQRWR	AKLSMRKCNADYLALRSSVLKVQAYRKATI	1140
Db	1174	AINIQQRWR	AKLSMRKCNADYLALRSSVLKVQAYRKATI	1233
Qy	1141	RLRAIMKMREQRENYLRLRNASILVQKRYRMRQMIQDRNAYL	LRTRKCIINVQRRWRATL	1200

Db 1293	1234	RLRAIMKMREQRENYLRLRNASILVQKRYRMRQQMIQDRNAYLRTRKCIINVQRRWRATL
Qy 1260	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER
Db 1353	1294	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER
Qy 1320	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRAKYRGTAQAVSCLQMHWNRHLLRKRENSFLQL
Db 1413	1354	QEYLHLREVTIKLQRRFHAQKSMRFMRAKYRGTAQAVSCLQMHWNRHLLRKRENSFLQL
Qy 1380	1321	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQVVLYQKQREAIK
Db 1473	1414	RQAAITLQRRYRARLNMIKQLKSYAQLKQAAITIQTRYRAKKAMQKQVVLYQKQREAIK
Qy 1440	1381	VQRRYRGNLEMRKQIEVYQKQRAVIRLQKWWRSIRDMLCKAGYRRIRLSSLSIQRKWR
Db 1533	1474	VQRRYRGNLEMRKQIEVYQKQRAVIRLQKWWRSIRDMLCKAGYRRIRLSSLSIQRKWR
Qy 1500	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML
Db 1593	1534	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML
Qy 1560	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQQKFRGKRLMIEQRNCF
Db 1653	1594	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQQKFRGKRLMIEQRNCF
Qy 1620	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ
Db 1713	1654	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ
Qy 1680	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLL
Db 1773	1714	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSRLDRLSRTVPHLL
Qy 1740	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM
Db 1833	1774	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM



Db	212	LKIQSAFRMAKAQKQFRLFKTAALVIQQNFRAWTAGRKQCMHEYIELRHAVLVLQSMWKGK	271
Qy 1105	1059	--RR-----MRSKWSMLEQRKEFQQLKRAAINIQQRWRAKLSMRKCNADYLALR	
Db	272	TLRRQLQRQHKAIIISYRMHVQQKKWKIMKKAALLIQKYRAYSIGREQNHLYLTK	331
Qy 1146	1106	SSVLKVQ-AYRKATIQMRI-----DRNHYYSLRKNVICLQQLRAIM	
Db	332	AAVVTLQSAYRGMKVRKRIKDCNKAAVTIQSKYRAYKTKKKYATYRASAIIRWYRGIK	391
Qy 1187	1147	KMREQRENYLRLRNASILVQKRY---RMRQ-----QMIQDRNAYLRTRK	
Db	392	ITNHQHKYENLKKTAIKIQSVYRGIRVRRHIQHMhraATFIKAMFKMHQSRI SYHTMRK	451
Qy 1231	1188	CIINVQRRWRATLQMRREKKNYLHLQTTTKRIQIKF---RAKREMKK-----	
Db	452	AAIVIQVRCRAYYQGKMQRKYLTILKAVKVLQASFRGVRVRRTLRKMQTAATLIQSNYR	511
Qy 1288	1232	---QRAEFLQLKKVTLVVQKRRRALLQMRKERQEYLHLREVTIKLQRRFHAQKSMRFMRA	
Db	512	RYRQQTYFNKLKKITKTQRYWAMKERNIQFQRYNKLHRSVIYIQAI FRGKKARRHLKM	571
Qy 1348	1289	KYRGTAAVSCLQMHWRNHLLRKRENSFLQLRQAATLQRRYRARLNMILKLSYAQLK	
Db	572	MH-----IAATLIQRRFRTLMRRR-----FLSLKKTALIQRYRAHL-CTKHHLQFLQVQ	622
Qy 1389	1349	QAATITQTRYR---AKKAMQKQ-----VVLYQKQREAI IKVQRRYRGNL	
Db	623	NAVIKIQSSYRRWMI RKRMRMHRAATFIQSTFRMHLHMRVQALKQASVVIQQQYQANR	682
Qy 1449	1390	EMRKQIEVYQKQKQAVIRLQKWWRSIRDMRLCKAGYRRIRLSSLSIQRKWRATVQARRQR	
Db	683	AAKLQRQHLYLRQHSVILQAAFRGMKTRRHLKS-----MHSSATLIQSRSRSLVRRR--	736
Qy 1509	1450	EIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAMLKARQDYQLI	
Db	737	--FISLKKATIFVQRKYRATICAKHKLYQFLHLRKAATIQSSYR-----RLMVKKKLQEM	790
Qy 1566	1510	QSSVILVQRKFRANRSMKQARQEFVQLRT---IAVHLQQKFRGKRLMIEQRNCFQLLRCS	
Db	791	QRAAVLIQATFRMHRT-----YITFQTWKHASILIQQHYRTYRAAKLQRE-----	835
Qy 1624	1567	MPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRR--RQKHQGLLD	

: ||| :| ||| ::| | :| ||| :

Db 836 -----NYIRQWHSAVVIQAAYKGMKARQLLREKHKASIV 869

Qy 1625 IRKRIAQLRQEAKAVNSVRC---KVQEAVRFLRGRF 1657

|: || | |:| | : :: ::

Db 870 IQSTYRMYRQ-----YCFYQKLQWATKIIQEKY 897

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2005, 17:10:37 ; Search time 8 Seconds  
(without alignments)  
262.817 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 8323 seqs, 1129788 residues

Total number of hits satisfying chosen parameters: 8323

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA New:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	140.5	1.5	895	7	US-11-150-406-2	Sequence 2, Appli
2	124.5	1.3	1463	1	US-10-971-982-3	Sequence 3, Appli
3	123	1.3	724	7	US-11-109-156-19	Sequence 19, Appl
4	122.5	1.3	1189	7	US-11-074-176-134	Sequence 134, App
5	113.5	1.2	1267	7	US-11-109-156-35	Sequence 35, Appl
6	103.5	1.1	828	1	US-10-501-039-2	Sequence 2, Appli
7	102	1.1	2516	1	US-10-647-956A-2	Sequence 2, Appli
8	100.5	1.1	648	7	US-11-109-156-17	Sequence 17, Appl



9	99.5	1.0	1420	7	US-11-077-550-110	Sequence 110, App
10	98	1.0	1087	7	US-11-102-978-2	Sequence 2, Appli
11	98	1.0	1138	1	US-10-509-422-4	Sequence 4, Appli
12	98	1.0	2897	1	US-10-499-715-2	Sequence 2, Appli
13	97.5	1.0	683	7	US-11-081-566-2	Sequence 2, Appli
14	97	1.0	1302	7	US-11-004-057-6	Sequence 6, Appli
15	97	1.0	3056	7	US-11-109-156-20	Sequence 20, Appl
16	95.5	1.0	605	1	US-10-689-742-140	Sequence 140, App
17	94.5	1.0	1493	7	US-11-004-057-21	Sequence 21, Appl
18	93	1.0	1311	1	US-10-509-422-5	Sequence 5, Appli
19	92.5	1.0	691	1	US-10-131-826A-16	Sequence 16, Appl
20	92	1.0	1137	1	US-10-499-715-4	Sequence 4, Appli
21	91.5	1.0	1315	7	US-11-077-550-141	Sequence 141, App
22	90.5	1.0	1618	1	US-10-984-645-2	Sequence 2, Appli
23	89.5	0.9	716	1	US-10-131-826A-96	Sequence 96, Appl
24	89.5	0.9	716	1	US-10-131-826A-512	Sequence 512, App
25	88.5	0.9	747	7	US-11-182-592-2	Sequence 2, Appli
26	88.5	0.9	1493	7	US-11-004-057-4	Sequence 4, Appli
27	85.5	0.9	903	1	US-10-689-742-142	Sequence 142, App
28	85.5	0.9	1168	1	US-10-509-422-2	Sequence 2, Appli
29	85	0.9	915	1	US-10-647-956A-6	Sequence 6, Appli
30	84	0.9	849	1	US-10-467-962B-53	Sequence 53, Appl
31	84	0.9	877	7	US-11-074-176-322	Sequence 322, App
32	84	0.9	883	7	US-11-074-176-88	Sequence 88, Appl
33	83.5	0.9	306	1	US-10-518-341-2	Sequence 2, Appli
34	83.5	0.9	478	1	US-10-689-742-184	Sequence 184, App
35	83.5	0.9	1663	1	US-10-982-545-6	Sequence 6, Appli
36	83	0.9	296	1	US-10-131-826A-26	Sequence 26, Appl
37	83	0.9	2053	7	US-11-013-759-9	Sequence 9, Appli
38	82.5	0.9	709	7	US-11-074-176-158	Sequence 158, App
39	82	0.9	437	1	US-10-967-648A-4	Sequence 4, Appli
40	81	0.9	457	1	US-10-131-826A-236	Sequence 236, App
41	81	0.9	511	7	US-11-012-762-66	Sequence 66, Appl
42	81	0.9	764	7	US-11-022-562-209	Sequence 209, App
43	81	0.9	867	1	US-10-131-826A-20	Sequence 20, Appl
44	80.5	0.8	439	7	US-11-034-569-16	Sequence 16, Appl
45	80.5	0.8	835	1	US-10-501-039-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
 US-11-150-406-2  
 ; Sequence 2, Application US/11150406  
 ; Publication No. US20050250164A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Muschler, John L  
 ; APPLICANT: Bissell, Mina J  
 ; TITLE OF INVENTION: Design of Novel Assays Based on the Newly Found Role of  
 ; TITLE OF INVENTION: Dystroglycan and Alpha-Dystroglycan Proteolysis in  
 Tumor Cell  
 ; TITLE OF INVENTION: Growth  
 ; FILE REFERENCE: IB-1398A  
 ; CURRENT APPLICATION NUMBER: US/11/150,406  
 ; CURRENT FILING DATE: 2005-06-09

```
; PRIOR APPLICATION NUMBER: 60/151,766
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 09/652,493
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 895
; TYPE: PRT
; ORGANISM: homo sapiens
US-11-150-406-2
```

```
Query Match          1.5%; Score 140.5; DB 7; Length 895;
Best Local Similarity 20.3%; Pred. No. 0.00067;
Matches 126; Conservative 84; Mismatches 252; Indels 159; Gaps
26;
```

```
Qy      26 RKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSP--TGAGKTMKSVVSAAVQQKKRMSAA 83
      | ::| ::| | : | : | | | | | | : | : | :
Db      207 RIDLLHRMRSFSEVELHNMKLVVNNRFLDMSAFMAGPGNPKKVVENGALLSWKLGCSL 266

Qy      84 AAPPSKQTRVVTAPSRPAAWA----HP-----PPQAPLVEKNVYKTP----- 121
      | ||:| | : :| | | | | : ::| |
Db      267 NQNSVPDIHGVEAPAREGAMSAQLGYPVVGWHIANKKPPLPKRVRRQIHATPTPVTAIGP 326

Qy     122 -----QEEPVIISPQPRS-----LKENLSP----MTPGNLL-----DVIDNLRFTPLT 160
      || | | | | | | :| | | | :| | :
Db     327 PTTAIQEPPSRIVPTPTSPAIAAPTETMAPPVDRDPVPGKPTVTIIRTRGAIQPTTLGPIQ 386

Qy     161 ETRGKGQATIFPDNLAAWPTPTLKGNVKSACA-----NDMRPRRITPDDLEDQPATNKT 213
      || | | | : || |:| |:| | :|| || :||| |
Db     387 PTRVSEAGTTVPQGIR--PTMTIPGYVEPTAVATPPTTTTKKPRVSTP-----KPATPST 439

Qy     214 FDVKHSETINISLDTLDCSRIDGQPHT--PLNKTTTIVHATHTRALACIHEEEGSPSP-- 269
      | | | :| | |: : || | | | | | | |
Db     440 -----DSTTTTTTRRPTKKPRTPRPVPRVTTKVSITRL-----ETASPPTR 479

Qy     270 -RTPTKSAIH-----DLKRDIKLVGSPLRKYSESMKDLSLLSPQTKYAIQGSMPNLN 320
      || | | :|| | | : : | | | : | | : : :
Db     480 IRTTTSGVPRGGEPNQRPELKNHIDRVDWVGTYFEVK-----IPSDTFYDHEDTTTDDKL 534

Qy     321 EMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQEFLFNHSEI--LAQSSRFNLH 378
      :: :: || : : :| : : :| | : | : : : :| :|
Db     535 KLTLKLREQQLVGEKSWVQFNSNSQLMYGLPDSSSHVGKHEYFMHATDKGGLSAVDAFEIH 594

Qy     379 EVGRKSVKGSPPV-----NPHKRRS--HELSEFSDAPSNESE----- 412
      | :| : : ||: : :|:|: | |
Db     595 VHRRPQGDRAPARFKAKFVGDPALVLNDIHKKIALVKKLAFAGDRNCSTITLQNITRGS 654

Qy     413 ----YRNETVAISPPKKQRV-----EDTTLPRSAAPANASA-----RSSSAHAWPHA 455
      : | |: : | |::| | | | | :| | :| :|
Db     655 IVVEWTNNTLPLEPCPKEQIAGLSRRIAEDDGKPR---PAFSNALEPDFKATSITVTGSG 711

Qy     456 QSKKFKLAQTMSLMKKP--ATPRKVRDTSIQPSVKLYDSELYMQTCINPDFAATTTIDP 513
      : : : : | | | :| | : | : :| :| | |
Db     712 SCRHLQFIPVVPVRRVPSEAPPTEVPDRDPE---KSEDDVYLHTVIPAVVVAAILLIAG 768
```

QY 514 FLASTMY-----LDEQA 525  
:| | |::||  
Db 769 IIAMICYRKKRKGLTLEDQA 789

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2005, 16:59:44 ; Search time 60 Seconds  
(without alignments)  
2984.323 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	9514	100.0	1861	2	T13845	microtubule-associ
2	361.5	3.8	1186	2	T19957	hypothetical prote
3	324.5	3.4	1088	2	T05464	hypothetical prote
4	324	3.4	1407	1	S28589	trichohyalin - rab
5	323.5	3.4	1549	1	A40691	trichohyalin - she
6	320	3.4	1830	1	S19188	myosin-V - chicken
7	314	3.3	1909	2	A45592	liver stage antige
8	304	3.2	1828	2	B59254	mysoin heavy chain
9	301	3.2	4574	2	G02520	plectin - human
10	301	3.2	4687	1	A39638	plectin - rat
11	300.5	3.2	1375	2	T05200	myosin heavy chain
12	300.5	3.2	1855	2	A59254	mysoin heavy chain
13	300	3.2	1374	2	D85390	myosin-like protei

14	294	3.1	4684	2	A59404	plectin [imported]
15	284.5	3.0	1805	1	A64224	hypothetical prote
16	277	2.9	1818	1	S73852	hypothetical prote
17	277	2.9	1898	1	A45973	trichohyalin - hum
18	271.5	2.9	1853	1	A46761	myosin heavy chain
19	267.5	2.8	2663	1	S28261	centromere protein
20	264	2.8	2442	2	T08621	centrosome associa
21	260	2.7	2245	2	T18278	myosin heavy chain
22	254.5	2.7	1837	2	T41023	probable nuclear p
23	252	2.6	1313	2	A48467	myosin heavy chain
24	251.5	2.6	1792	2	T13939	myosin V - fruit f
25	250	2.6	2101	2	A42184	nuclear mitotic ap
26	246.5	2.6	2022	2	T43214	ovtl protein - nem
27	244	2.6	1846	2	A59289	myr 6, unconventio
28	243.5	2.6	812	2	A53016	myosin heavy chain
29	243.5	2.6	1410	1	A57013	early endosome ant
30	239.5	2.5	1940	2	A59287	myosin heavy chain
31	235	2.5	2253	2	T30336	nuclear/mitotic ap
32	235	2.5	3187	2	JC5837	364K Golgi complex
33	233.5	2.5	1643	2	T07961	myosin heavy chain
34	231	2.4	968	2	T00322	hypothetical prote
35	230.5	2.4	1211	2	C83110	probable exonuclea
36	228.5	2.4	1679	2	S48385	hypothetical prote
37	228	2.4	1957	2	A59294	skeletal myosin -
38	227.5	2.4	3259	1	A56539	giantin - human
39	227	2.4	1957	2	A45627	myosin heavy chain
40	227	2.4	3678	2	S28916	dystrophin - mouse
41	226	2.4	2346	2	T13829	Tpr homolog - frui
42	226	2.4	2626	2	T31099	myosin-RhoGAP prot
43	225.5	2.4	3225	2	I52300	giantin - human
44	225	2.4	3685	1	A27605	dystrophin, muscle
45	224.5	2.4	1927	2	A59236	embryonic muscle m

# ALIGNMENTS

## RESULT 1

T13845

microtubule-associated protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T13845

R;Saunders, R.D.; Avides, M.C.; Howard, T.; Gonzalez, C.; Glover, D.M.

J. Cell Biol. 137, 881-890, 1997

A;Title: The *Drosophila* gene abnormal spindle encodes a microtubule-associated protein that associates with the polar regions of the mitotic spindle.

A;Reference number: Z17792; MUID:97296495; PMID:9151690

A;Accession: T13845

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1861 <SAU>

A;Cross-references: UNIPROT:O01401; UNIPARC:UPI000007E201; EMBL:U95171;

NID:g1930121; PID:g1930122; PIDN:AAB51540.1

C;Genetics:

A;Gene: asp

A;Cross-references: FlyBase:FBgn0000140

C;Function:

A;Description: is required for the normal function of the mitotic spindle

Query Match 100.0%; Score 9514; DB 2; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSPT	60
Db	1	MELVWSPVLEVACKETLQLIDNRNFRKEVMIILKSKSNQPVKNPRKFPTVGKTLQLKSPT	60
Qy	61	GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRVVTAPSRPAAWAHPPPQAPLVEKNVYKT	120
Db	61	GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRVVTAPSRPAAWAHPPPQAPLVEKNVYKT	120
Qy	121	PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT	180
Db	121	PQEEPVIYISPPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT	180
Qy	181	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTDFVKHSETINISLDTLDCSRIDGQPHT	240
Db	181	PTLKGNVKSCANDMRPRRITPDDLEDQPATNKTDFVKHSETINISLDTLDCSRIDGQPHT	240
Qy	241	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPKSAIHDLKRDIKLVGSPLRKYSESMKD	300
Db	241	PLNKTTTIVHATHTRALACIHEEEGPSPPRTPKSAIHDLKRDIKLVGSPLRKYSESMKD	300
Qy	301	LSLLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	360
Db	301	LSLLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIQIKAKDLNSSSSSEASLAGQQE	360
Qy	361	FLFNHSEILAQSSRFLNHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI	420
Db	361	FLFNHSEILAQSSRFLNHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI	420
Qy	421	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	480
Db	421	SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKVRD	480
Qy	481	TSIQPSVKLYDSELYMQTCINPDPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL	540
Db	481	TSIQPSVKLYDSELYMQTCINPDPFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL	540
Qy	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF	600
Db	541	VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF	600
Qy	601	SEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLRGLGVVFE	660
Db	601	SEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLRGLGVVFE	660
Qy	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSQKILFLLPFL	720
Db	661	KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSQKILFLLPFL	720
Qy	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRRLGTVLQHRQTF	780

Db	721	DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSESELLANIGDITRELRRLGTVLQHRQTF	780
Qy	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPASRLQRI FNVKLALGALG	840
Db	781	LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPASRLQRI FNVKLALGALG	840
Qy	841	EANFQLGGDIAAQDIVDGHREKTL SLLWQLIYKFRSPKFHAAATVLQKWRRHWLHVVIQ	900
Db	841	EANFQLGGDIAAQDIVDGHREKTL SLLWQLIYKFRSPKFHAAATVLQKWRRHWLHVVIQ	900
Qy	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAIILQKFTRRYLAQKQLYQ	960
Db	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAIILQKFTRRYLAQKQLYQ	960
Qy	961	SYHSIITIQRWWRAQQQLGRQHRQRFVELREAAIFLQRIWRRRLF AAKLLAAAETARLQRS	1020
Db	961	SYHSIITIQRWWRAQQQLGRQHRQRFVELREAAIFLQRIWRRRLF AAKLLAAAETARLQRS	1020
Qy	1021	QKQQAASYSIQMWRTYQLGRIQRHEFLRQRD LIMFVQRRMRSKWSMLEQRKEFQQLKRA	1080
Db	1021	QKQQAASYSIQMWRTYQLGRIQRHEFLRQRD LIMFVQRRMRSKWSMLEQRKEFQQLKRA	1080
Qy	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1140
Db	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1140
Qy	1141	RLRAIMKMREQRENYLRLRNAS ILVQKRYMRQQMIQDRNAYLRTRKCI INVQRRWRATL	1200
Db	1141	RLRAIMKMREQRENYLRLRNAS ILVQKRYMRQQMIQDRNAYLRTRKCI INVQRRWRATL	1200
Qy	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1260
Db	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1260
Qy	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTOAAVSC LQMHWRNHLLRKRENSFLQL	1320
Db	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTOAAVSC LQMHWRNHLLRKRENSFLQL	1320
Qy	1321	RQAAITLQRRYRARLNMIKQLKS YAQLKQAAITIQTRYRAKKAMQKQVVLYQKQREAIK	1380
Db	1321	RQAAITLQRRYRARLNMIKQLKS YAQLKQAAITIQTRYRAKKAMQKQVVLYQKQREAIK	1380
Qy	1381	VQRRYRGNLEMRKQIEVYQKQRAVIRLQKWWSIRDMLCKAGYRRIRLSSLSIQRKWR	1440
Db	1381	VQRRYRGNLEMRKQIEVYQKQRAVIRLQKWWSIRDMLCKAGYRRIRLSSLSIQRKWR	1440
Qy	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML	1500
Db	1441	ATVQARRQREIFLSTIRKVRMLQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAML	1500
Qy	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQ QKFRGKRLMIEQRNCF	1560
Db	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQ QKFRGKRLMIEQRNCF	1560
Qy	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1620
Db	1561	QLLRCSMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1620

Qy	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLL	1680
Db	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLL	1680
Qy	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1740
Db	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1740
Qy	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ	1800
Db	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ	1800
Qy	1801	NARKPPPMSTSGRYKSQKINFPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM	1860
Db	1801	NARKPPPMSTSGRYKSQKINFPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM	1860
Qy	1861	F 1861	
Db	1861	F 1861	

RESULT 2

T19957

hypothetical protein C45G3.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T19957

R;Barlow, K.

submitted to the EMBL Data Library, March 1997

A;Reference number: Z19203

A;Accession: T19957

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1186 <WIL>

A;Cross-references: UNIPROT:O17666; UNIPARC:UPI000007D3B1; EMBL:Z92780;

PIDN:CAB07174.1; GSPDB:GN00019; CESP:C45G3.1

A;Experimental source: clone C45G3

C;Genetics:

A;Gene: CESP:C45G3.1

A;Map position: 1

A;Introns: 21/1; 255/2; 363/2; 575/3; 893/3; 1017/2; 1042/1

C;Superfamily: *Caenorhabditis elegans* hypothetical protein C45G3.1

Query Match 3.8%; Score 361.5; DB 2; Length 1186;

Best Local Similarity 18.6%; Pred. No. 5.1e-12;

Matches 244; Conservative 216; Mismatches 480; Indels 375; Gaps 47;

Qy	458	KKFKLAQTMslmkk---PATPRKVRDTsiQPSVKLYDSELYMQTCINPDpFAATTTIDPF	514
		:                :  : :  :  :   :: :	
Db	21	EKRLLDQVKSNTKKIDLRATERAFLESS-----PTSMNMRTPLNPS-ISSSVSDSPI	71
Qy	515	LASTMYLDEQAVDRHQADfKKWLNALVSIPADLDADLNNKIDVGKLFNEV-----	564
		:   :  ::     : ::  :  :: :	
Db	72	LS----YDEKA-NKQIIALATWCNTM-----MELDVSEEMDLGESKAECRNiQKMLKK	120
Qy	565	RNkELVVAPTKEEQSMNY---LTKYRLETLRKAaVELFFSEQMRLPCSKVAVYVnKQALR	621



Db	121	RS	DTSE	VENTQ	ENARR	RYQ	RIFE	KN	DP	EV	VKK	CK	Q	L	DD	S	G	M	D	---	AS	I	K	D	L	L	S	K	N	N	V	A	177																													
Qy	622	IR	SD	R	N	L	H	L	D	V	M	Q	R	T	I	E	L	L	L	C	F	N	P	L	W	R	L	G	L	E	V	V	F	G	E	K	I	Q	M	Q	S	N	R	D	I	V	G	L	S	T	F	I	L	N	R	L	681					
Db	178	IR	KE	H	A	V	Y	N	D	I	G	L	Q	T	T	L	L	H	T	F	L	S	F	H	P	A	W	L	K	T	A	E	A	I	F	N	T	R	I	D	A	Q	P	K	H	L	M	K	K	L	S	Q	F	F	L	D	L	V	237			
Qy	682	FR	N	---	K	C	E	E	Q	R	Y	S	K	A	Y	---	T	L	T	E	E	Y	A	E	T	I	K	H	S	L	Q	K	I	L	F	L	P	F	L	D	Q	A	K	Q	K	R	I	V	K	H	N	P	C	L	F	736						
Db	238	FS	N	P	T	M	L	K	N	K	F	A	Q	G	S	G	K	P	I	T	E	A	G	E	A	L	H	K	H	F	L	S	V	S	M	K	L	M	F	L	I	E	T	A	H	T	H	R	V	I	P	N	L	T	R	I	F	297				
Qy	737	V	K	S	P	H	K	E	T	K	D	I	L	L	R	F	S	S	E	L	L	-	A	N	I	G	D	I	T	R	E	L	R	R	L	G	Y	V	L	Q	H	R	Q	T	F	L	D	E	F	D	Y	A	F	N	N	L	A	V	D	L	795	
Db	298	T	K	S	S	H	F	N	C	L	D	D	V	F	S	E	L	T	K	E	L	T	G	S	S	A	T	F	K	K	A	F	A	K	V	G	F	I	P	T	Y	R	Q	S	F	I	E	N	Y	D	Y	Q	A	K	G	F	S	-	D	F	356	
Qy	796	R	D	G	V	R	L	T	R	V	E	V	I	---	L	L	R	D	D	L	T	R	Q	L	R	V	P	A	I	S	R	L	Q	R	I	F	N	V	K	L	A	L	G	A	L	G	E	A	N	F	Q	L	G	---	G	D	I	850				
Db	357	S	D	G	L	I	L	A	K	L	L	E	T	V	G	E	M	P	H	G	Q	I	L	L	K	R	D	P	A	G	D	R	I	R	K	I	G	N	V	K	I	V	L	Q	E	M	S	---	S	L	G	V	P	T	D	N	V	412				
Qy	851	A	A	Q	D	I	V	D	G	H	R	E	K	T	L	S	L	L	W	Q	L	I	Y	K	F	R	S	P	K	F	H	A	A	T	V	L	Q	K	W	R	R	H	L	H	V	I	Q	R	R	I	R	---	H	905								
Db	413	N	A	E	S	I	V	G	G	K	D	E	I	L	S	I	L	W	A	I	I	-----	G	V	R	V	A	K	E	Q	R	I	K	V	T	R	V	S	E	450																						
Qy	906	K	E	L	M	R	R	H	R	A	A	-----	T	V	I	Q	A	V	F	R	G	H	Q	M	R	K	Y	V	K	L	F	K	T	E	R	T	Q	A	A	I	L	Q	K	F	T	R	R	Y	L	A	Q	K	Q	957								
Db	451	E	R	T	P	K	R	R	S	A	V	H	D	D	M	S	S	E	V	L	K	M	C	K	I	Y	G	R	Q	M	E	---	I	E	V	M	D	L	D	S	L	S	D	G	C	L	L	A	K	L	W	T	T	F	G	T	N	S	T	508		
Qy	958	L	Y	Q	S	Y	H	S	I	I	T	I	Q	R	W	W	R	A	Q	Q	L	G	R	Q	H	R	Q	R	F	V	E	L	R	E	A	A	I	F	L	Q	R	---	I	W	R	R	R	L	F	A	K	L	L	A	A	E	T	A	1015			
Db	509	P	I	Q	D	Y	D	G	---	L	S	L	W	-----	E	K	V	S	V	A	E	L	E	L	C	I	Q	R	G	L	D	Q	N	M	A	L	F	V	K	M	F	L	-----	E	549																	
Qy	1016	R	L	Q	R	S	Q	K	Q	Q	A	A	S	Y	I	Q	M	Q	W	R	T	Y	Q	L	G	R	I	Q	R	H	E	-----	F	L	-----	R	Q	R	D	L	I	M	F	---	1056																	
Db	550	R	L	G	M	I	Q	D	L	N	E	K	A	T	K	I	Q	R	M	W	K	A	Y	----	V	Q	R	K	N	T	P	K	L	Y	F	I	V	Q	Q	L	L	A	D	S	S	I	P	R	N	S	V	S	P	F	S	N	N	605				
Qy	1057	----	V	Q	R	R	M	R	S	K	W	S	M	L	E	Q	R	K	E	F	Q	Q	L	K	R	A	A	I	N	I	Q	Q	R	W	R	A	K	L	S	M	R	K	C	N	A	D	Y	L	A	L	R	S	S	V	L	K	V	Q	1112			
Db	606	V	T	F	T	V	P	R	T	P	R	N	---	N	I	L	T	E	R	P	S	L	S	Q	I	P	S	S	-----	R	Q	S	M	D	S	T	F	N	D	A	T	F	T	V	S	R	D	S	I	E	S	M	N	655								
Qy	1113	A	Y	R	K	A	-----	T	I	Q	M	R	I	D	R	N	H	Y	S	L	R	K	N	V	I	C	L	Q	Q	R	L	R	A	I	M	K	M	R	E	Q	R	E	N	Y	L	R	L	R	N	A	S	I	1163									
Db	656	K	M	Q	K	T	P	L	R	G	T	F	T	R	K	T	I	A	M	V	I	E	E	D	D	S	E	N	N	E	T	V	V	P	S	T	L	K	K	R	T	V	V	R	M	E	H	N	A	E	V	F	-----	707								
Qy	1164	L	V	Q	K	R	Y	R	M	R	Q	Q	M	I	Q	D	R	N	A	Y	L	R	T	R	K	C	I	I	N	V	Q	R	R	W	R	A	T	L	Q	M	R	R	E	R	K	N	Y	L	H	L	Q	T	T	T	K	R	I	Q	I	K	F	1223
Db	708	-----	R	E	Q	D	E	D	D	E	N	-----	Q	D	K	D	T	V	A	P	S	A	E	-----	N	L	D	S	P	P	S	D	I	P	L	E	T	740																								
Qy	1224	R	A	K	R	E	M	K	K	Q	R	A	E	F	L	Q	L	K	K	V	T	L	V	V	Q	K	R	R	R	A	L	L	Q	M	R	K	E	R	Q	E	Y	L	H	L	R	E	V	T	I	K	L	Q	R	R	F	H	A	Q	K	S	M	1283
Db	741	L	S	S	I	P	S	A	S	Q	S	A	I	F	L	Q	D	S	E	-----	T	G	K	E	M	763																																				
Qy	1284	R	F	M	R	A	K	Y	R	G	T	---	Q	A	A	V	S	C	L	Q	M	H	R	N	H	L	L	R	K	R	E	R	N	S	F	L	Q	L	R	Q	A	A	I	T	L	Q	R	R	Y	R	A	R	L	N	M	I	K	Q	1340			
Db	764	H	V	P	K	A	E	D	V	G	V	V	L	E	A	S	D	S	P	V	A	L	E	G	N	-----	E	A	S	Y	D	G	Q	K	I	E	N	L	E	T	F	E	I	K	E	808																
Qy	1341	L	K	S	Y	A	Q	L	K	Q	A	A	I	T	I	Q	T	R	Y	R	A	K	K	A	M	Q	K	Q	V	V	L	Y	Q	K	Q	R	E	A	I	K	V	Q	R	R	Y	R	G	N	L	E	M	R	K	Q	I	E	V	Y	Q	K	1400	

Db	809	GKTQEDLPSKSPMDPTQTSGSPLVEFRMTEEQERLEMLFQSLSEDQKNFVKTNLNSVSIE	868
Qy	1401	QRQAVIRLQKWWSIRDMLCKAGYRRIRLSSLSIQRKWRATVQARRQREIFLSTIRKVR	1460
		::    :::  ::	
Db	869	DDANTPELRRILRQTRELK-----RKQOEI-----AR	895
Qy	1461	LMQAFIRATLLMRQQRREFEMKRRAAVVIQRRFRARCAMLKARQDYQLIQSSVILVQRKF	1520
		:     :  :     :      ::   :	
Db	896	KLGNIERNALAVRDGGEDSSDSRSDA-----GHDVAILHGDDSQLFENSMQLDQK--	945
Qy	1521	RANRSMKQARQEFVQLRTIAVHLQQKFRGKRLMIEQRNCFQLLRCSMPGFAQARARGFMAR	1580
Db	946	-----	945
Qy	1581	KRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQGLLDIRKRIAQ---LRQEA	1636
		:::    :   ::   : : : :     :	
Db	946	---SQLQNDDETQILENKKKAAVVIQKMIIRGFIARRKFQME-ISNIRNRMIQYNHILAQED	1001
Qy	1637	KAV-----NSVRCKVQEA VRFLRGRFIASDALAVL-----SQLDRLSRTVPHLL	1680
		: :      ::     : :       : :  ::	
Db	1002	EQIGIEEMEDKSVEAKLKKCA--LHG--LTNDNLHVHVHAATVIDRVTDLVPSLL	1052

us-09-914-698-1.rup

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2005, 16:42:06 ; Search time 301 Seconds  
(without alignments)  
4362.087 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9482	99.7	1954	1	ASP_DROME	Q9vc45 drosophila
2	3093	32.5	1399	2	Q7QAG9_ANOGA	Q7qag9 anopheles g
3	1407.5	14.8	3452	1	ASPM_CANFA	P62286 canis fami
4	1392	14.6	3371	1	ASPM_BOVIN	P62285 bos taurus
5	1385	14.6	3469	1	ASPM_SAIIB	P62296 saimiri bol
6	1374	14.4	3122	1	ASPM_MOUSE	Q8cj27 mus musculu
7	1373	14.4	3477	1	ASPM_PANTR	P62293 pan troglod
8	1366.5	14.4	3476	1	ASPM_GORGO	P62289 gorilla gor
9	1363	14.3	3471	1	ASPM_PONPY	P62294 pongo pygma
10	1362.5	14.3	3473	1	ASPM_AOTVO	P62283 aotus vocif
11	1360	14.3	3477	1	ASPM_HUMAN	Q8izt6 homo sapien
12	1360	14.3	3477	1	ASPM_HYLLA	P62290 hylobates l
13	1360	14.3	3477	2	Q5VYL3_HUMAN	Q5vy13 homo sapien
14	1359	14.3	3374	1	ASPM_SHEEP	P62297 ovis aries
15	1356.5	14.3	3461	1	ASPM_FELCA	P62288 felis silve
16	1353.5	14.2	3475	2	Q69AX5_CERAE	Q69ax5 cercopithec
17	1348.5	14.2	3479	1	ASPM_MACMU	P62292 macaca mula
18	1343	14.1	3476	1	ASPM_MACFA	P62291 macaca fasc
19	1330.5	14.0	3477	1	ASPM_COLGU	P62287 colobus gue
20	1321.5	13.9	2220	2	Q4S6P6_TETNG	Q4s6p6 tetraodon n
21	1008	10.6	1857	2	Q4G1G9_MOUSE	Q4g1g9 mus musculu

us-09-914-698-1.rup							
22	999.5	10.5	1892	2	Q4G1H1_HUMAN	Q4g1h1	homo sapien
23	871	9.2	1142	2	Q5VYL4_HUMAN	Q5vyl4	homo sapien
24	744.5	7.8	1564	1	ASPM_ATEGE	P62284	ateles geof
25	729.5	7.7	1527	1	ASPM_SAGLB	P62295	saguinus la
26	710.5	7.5	1389	2	Q4G1H0_HUMAN	Q4g1h0	homo sapien
27	547.5	5.8	1062	2	Q4G1H2_HUMAN	Q4g1h2	homo sapien
28	414	4.4	1275	2	Q61KL0_CAEBR	Q61kl0	caenorhabdi
29	396.5	4.2	981	2	Q5BGB6_EMENI	Q5bgb6	aspergillus
30	374	3.9	1761	2	Q4QC85_LEIMA	Q4qc85	leishmania
31	370.5	3.9	1110	2	Q84Z28_ORYSA	Q84z28	oryza sativ
32	363.5	3.8	1015	2	Q4WJS9_ASPFU	Q4wjs9	aspergillus
33	361.5	3.8	1186	2	O17666_CAEEL	O17666	caenorhabdi
34	324.5	3.4	1088	2	O49705_ARATH	O49705	arabidopsis
35	324	3.4	1407	1	TRHY_RABIT	P37709	oryctolagus
36	323.5	3.4	1549	1	TRHY_SHEEP	P22793	ovis aries
37	317.5	3.3	1829	1	MYO5A_CHICK	Q02440	gallus gall
38	314	3.3	1909	2	Q25893_PLAFA	Q25893	plasmodium
39	303.5	3.2	2779	1	LVA_DROME	Q8mss1	drosophila
40	302.5	3.2	1851	2	Q4LE86_PIG	Q4le86	sus scrofa
41	301.5	3.2	1855	1	MYO5A_HUMAN	Q9y4i1	homo sapien
42	301	3.2	4515	2	Q6S379_HUMAN	Q6s379	homo sapien
43	301	3.2	4525	2	Q6S381_HUMAN	Q6s381	homo sapien
44	301	3.2	4533	2	Q6S382_HUMAN	Q6s382	homo sapien
45	301	3.2	4547	2	Q6S376_HUMAN	Q6s376	homo sapien

#### ALIGNMENTS

#### RESULT 1

##### ASP\_DROME

ID ASP\_DROME STANDARD; PRT; 1954 AA.  
AC Q9VC45; O01401; Q8SX66;  
DT 25-OCT-2004 (Rel. 45, Created)  
DT 25-OCT-2004 (Rel. 45, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Abnormal spindle protein.  
GN Name=asp; ORFNames=CG6875;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephynroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

us-09-914-698-1.rup

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE OF 94-1954, FUNCTION, SUBCELLULAR LOCATION, AND  
RP DEVELOPMENTAL STAGE.  
RC STRAIN=Oregon-R;  
RX PubMed=9151690; DOI=10.1083/jcb.137.4.881;  
RA Saunders R.D.C., do Carmo Avides M., Howard T.I.A., Gonzalez C.,  
RA Glover D.M.;  
RT "The *Drosophila* gene abnormal spindle encodes a novel microtubule-  
RT associated protein that associates with the polar regions of the  
RT mitotic spindle.";  
RL J. Cell Biol. 137:881-890(1997).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 700-1954.  
RC STRAIN=Berkeley; TISSUE=Embryo;  
RX MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A *Drosophila* full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
RN [5]  
RP FUNCTION, AND SUBCELLULAR LOCATION.  
RX PubMed=10073938; DOI=10.1126/science.283.5408.1733;  
RA do Carmo Avides M., Glover D.M.;  
RT "Abnormal spindle protein, Asp, and the integrity of mitotic  
RT centrosomal microtubule organizing centers.";  
RL Science 283:1733-1735(1999).  
RN [6]

RP FUNCTION.  
RX PubMed=15242765; DOI=10.1016/j.yexcr.2004.03.054;  
RA Riparbelli M.G., Massarelli C., Robbins L.G., Callaini G.;  
RT "The abnormal spindle protein is required for germ cell mitosis and  
RT oocyte differentiation during *Drosophila* oogenesis.";  
RL Exp. Cell Res. 298:96-106(2004).  
CC -!- FUNCTION: Required to maintain the structure of the centrosomal  
CC microtubule-organizing center (MTOC) during mitosis. May have a  
CC preferential role in regulating neurogenesis. Required for germ  
CC cell mitosis and oocyte differentiation.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear. During interphase  
CC in syncytial embryos distribution is cytoplasmic. On entering  
CC mitosis, moves to polar regions of the spindle immediately  
CC surrounding the centrosome. At telophase, migrates to microtubules  
CC on the spindle side of both daughter nuclei. The nuclear-  
CC cytoplasmic distribution could be regulated by the availability of  
CC calmodulin.  
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically in  
CC embryos.  
CC -!- SIMILARITY: Contains 1 CH (calponin-homology) domain.  
CC -!- SIMILARITY: Contains 5 IQ domains.  
CC -!- CAUTION: Ref.4 sequence differs from that shown due to intron  
CC retention.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
DR EMBL; AE003749; AAF56330.3; -; Genomic\_DNA.  
DR EMBL; U95171; AAB51540.1; -; mRNA.  
DR EMBL; AY094825; AAM11178.1; ALT\_SEQ; mRNA.  
DR PIR; T13845; T13845.  
DR Ensembl; CG6875; *Drosophila melanogaster*.  
DR FlyBase; FBgn0000140; asp.  
DR GO; GO:0005875; C:microtubule associated complex; IDA.  
DR GO; GO:0005815; C:microtubule organizing center; IDA.  
DR GO; GO:0008017; F:microtubule binding; IDA.  
DR GO; GO:0004672; F:protein kinase activity; IDA.  
DR GO; GO:0000226; P:microtubule cytoskeleton organization and b. . .; IDA.  
DR InterPro; IPR001715; Calponin\_act\_bd.  
DR InterPro; IPR000048; IQ\_Cam\_bd\_region.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00612; IQ; 19.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00015; IQ; 16.  
DR PROSITE; PS50021; CH; 1.  
DR PROSITE; PS50096; IQ; 5.  
KW Calmodulin-binding; Cell cycle; Cell division; Coiled coil;  
KW Developmental protein; Differentiation; Microtubule; Mitosis;  
KW Nuclear protein; Oogenesis; Repeat.  
FT DOMAIN 836 968 CH.  
FT DOMAIN 1004 1033 IQ 1.  
FT DOMAIN 1386 1415 IQ 2.  
FT DOMAIN 1467 1496 IQ 3.  
FT DOMAIN 1656 1687 IQ 4.  
FT DOMAIN 1690 1721 IQ 5.  
FT COILED 1614 1641 Potential.  
FT COMPBias 1063 1749 Arg-rich.  
FT CONFLICT 811 811 L -> P (in Ref. 3).  
FT CONFLICT 898 898 M -> V (in Ref. 3).  
FT CONFLICT 1129 1129 S -> T (in Ref. 3).

```

us-09-914-698-1.rup
FT CONFLICT 1138 1139 QQ -> HE (in Ref. 3).
FT CONFLICT 1761 1761 R -> Q (in Ref. 1).
SQ SEQUENCE 1954 AA; 230180 MW; 4912B4E20CA9E659 CRC64;

```

```

Query Match          99.7%; Score 9482; DB 1; Length 1954;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1855; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 MELVWSPVLEVACKETLQLIDNRNFRKEVMILKSKSNQPVKNPRKFPTVGKTLQLKSPT 60
Db      94 MELVWSPVLEVACKETLQLIDNRNFRKEVMILKSKSNQPVKNPRKFPTVGKTLQLKSPT 153
Qy     61 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRWVTAPSRPAAWAHPPPQAPLVEKNVYKT 120
Db    154 GAGKTMKSVVSAAVQQKKRMSAAAAPPSKQTRWVTAPSRPAAWAHPPPQAPLVEKNVYKT 213
Qy    121 PQEEPVIYISPQPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT 180
Db    214 PQEEPVIYISPQPRSLKENLSPMTPGNLLDVIDNLRFTPLTETRGKGQATIFPDNLAAWPT 273
Qy    181 PTLKGNVSKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPH 240
Db    274 PTLKGNVSKSCANDMRPRRITPDDLEDQPATNKTFDVKHSETINISLDTLDCSRIDGQPH 333
Qy    241 PLNKTTHIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSSEMKD 300
Db    334 PLNKTTHIVHATHTRALACIHEEEGPSPPRTPTKSAIHDLKRDIKLVGSPLRKYSSEMKD 393
Qy    301 LSLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIIKAKDLNSSSSSEASLAGQQE 360
Db    394 LSLSPQTKYAIQGSMPNLNEMKIRSIEQNRYYQEQQIIKAKDLNSSSSSEASLAGQQE 453
Qy    361 FLFNHSEILAQSSRFNLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI 420
Db    454 FLFNHSEILAQSSRFNLHEVGRKSVKGSVPKNPHKRRSHELFSFDAPSNESELYRNETVAI 513
Qy    421 SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKV 480
Db    514 SPPKKQRVEDTTLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTMSLMKKPATPRKV 573
Qy    481 TSIQPSVKLYDSELYMQTCINPDFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL 540
Db    574 TSIQPSVKLYDSELYMQTCINPDFAATTTIDPFLASTMYLDEQAVDRHQADFKKWLNAL 633
Qy    541 VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF 600
Db    634 VSIPADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRLETLRKAAVELFF 693
Qy    601 SEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLRLGLEVVFG 660
Db    694 SEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLLCFNPLWLRLGLEVVFG 753
Qy    661 KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSQKILFLLPFL 720
Db    754 KIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKHSQKILFLLPFL 813
Qy    721 DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRLGYVLQHRQTF 780
Db    814 DQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRLGYVLQHRQTF 873
Qy    781 LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPAISRLQRIFNVKLALGALG 840
Db    874 LDEFDYAFNNLAVDLRDGVRLTRVMEVILLRDDLTRQLRVPAISRLQRIFNVKLALGALG 933

```

Qy	841	EANFQLGGDIAAQDIVDGHREKTLSELLWQLIYKFRSPKFHAAATVLQKWRRHWLHVVIQ	900
Db	934	EANFQLGGDIAAQDIVDGHREKTLSELLWQLIYKFRSPKFHAAATVLQKWRRHWLHVVIQ	993
Qy	901	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKL FK TERTQAAIILQKFTRRYLAQKQLYQ	960
Db	994	RRIRHKELMRRHRAATVIQAVFRGHQMRKYVKL FK TERTQAAIILQKFTRRYLAQKQLYQ	1053
Qy	961	SYHSIITIQRWWRAQQLGRQHRQRFVELREAAIFLQRIWRRRLFAKKLLAAAE TARLQRS	1020
Db	1054	SYHSIITIQRWWRAQQLGRQHRQRFVELREAAIFLQRIWRRRLFAKKLLAAAE TARLQRS	1113
Qy	1021	QKQQAASYSIQMWRTYQLGRIQRHEFLRQRDLIMFVQRRMRSKWSMLEQRKEFQQLKRA	1080
Db	1114	QKQQAASYSIQMWRSYQLGRIQRQQLRQRDLIMFVQRRMRSKWSMLEQRKEFQQLKRA	1173
Qy	1081	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1140
Db	1174	AINIQQRWRAKLSMRKCNADYLALRSSVLKVQAYRKATI QMRIDRNHYYS LRKNVICLQQ	1233
Qy	1141	RLRAIMKMREQRENYLRLRNASILVQKRYMRQMQMIQDRNAYLRTRKCIINVQRRWRATL	1200
Db	1234	RLRAIMKMREQRENYLRLRNASILVQKRYMRQMQMIQDRNAYLRTRKCIINVQRRWRATL	1293
Qy	1201	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1260
Db	1294	QMRREKKNYLHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKER	1353
Qy	1261	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTA AVSCLQMHWNRHLLRKRENSFLQL	1320
Db	1354	QEYLHLREVTIKLQRRFHAQKSMRFMRACYRGTA AVSCLQMHWNRHLLRKRENSFLQL	1413
Qy	1321	RQAAITLQRRYRARLNM IKQLKSYAQLKQAAITIQTRYAKKAMQKQVVL YQKQREAIK	1380
Db	1414	RQAAITLQRRYRARLNM IKQLKSYAQLKQAAITIQTRYAKKAMQKQVVL YQKQREAIK	1473
Qy	1381	VQRRYRG NLEMRKQIEVYQKQRQAVIRLQKWWRSIRDMLCKAGYRRIRLSSLSIQRKWR	1440
Db	1474	VQRRYRG NLEMRKQIEVYQKQRQAVIRLQKWWRSIRDMLCKAGYRRIRLSSLSIQRKWR	1533
Qy	1441	ATVQARRQREIFLSTIRKVRLMQAFIRATLLMRQRRREFEMKRRAAVVIQRRFRARCAML	1500
Db	1534	ATVQARRQREIFLSTIRKVRLMQAFIRATLLMRQRRREFEMKRRAAVVIQRRFRARCAML	1593
Qy	1501	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQQKFRGKRLMIEQRNCF	1560
Db	1594	KARQDYQLIQSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQQKFRGKRLMIEQRNCF	1653
Qy	1561	QLLRCSMPGFQARARGFMARKRFQALMTPPEMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1620
Db	1654	QLLRCSMPGFQARARGFMARKRFQALMTPPEMDLIRQKRAAKVIQRYWRGYLIRRRQKHQ	1713
Qy	1621	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSQLDRLSRTVP HLL	1680
Db	1714	GLLDIRKRIAQLRQEAKAVNSVRCKVQEAVRFLRGRFIASDALAVLSRLDRLSRTVP HLL	1773
Qy	1681	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1740
Db	1774	MWCSEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQM	1833
Qy	1741	LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ	1800



```

                                us-09-914-698-1.rup
Db      1834 LLRWCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQ 1893
Qy      1801 NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM 1860
      |||
Db      1894 NARKPPPMSTSGRYKSQKINFTPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCKLQIDM 1953
Qy      1861 F 1861
      |
Db      1954 F 1954

```

## RESULT 2

Q7QAG9\_ANOGA

```

ID      Q7QAG9_ANOGA PRELIMINARY;          PRT; 1399 AA.
AC      Q7QAG9;
DT      01-MAR-2004 (TrEMBLrel. 26, Created)
DT      01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      ENSANGP00000021262 (Fragment).
GN      ORFNames=ENSANGG00000018773;
OS      Anopheles gambiae str. PEST.
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC      Anophelinae; Anopheles.
OX      NCBI_TaxID=180454;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=PEST;
RG      The Anopheles gambiae Sequence Committee;
RT      "Anopheles gambiae re-annotation.";
RL      Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=PEST;
RG      The Anopheles gambiae Sequence Committee;
RL      Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
CC      -!- CAUTION: The sequence shown here is derived from an
CC      EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC      preliminary data.
DR      EMBL; AAAB01008888; EAA08939.2; -; Genomic_DNA.
DR      InterPro; IPR001715; Calponin-like.
DR      InterPro; IPR000048; IQ_region.
DR      Pfam; PF00307; CH; 1.
DR      Pfam; PF00612; IQ; 26.
DR      SMART; SM00033; CH; 1.
DR      SMART; SM00015; IQ; 19.
DR      PROSITE; PS50021; CH; 1.
DR      PROSITE; PS50096; IQ; 5.
FT      NON_TER      1      1
SQ      SEQUENCE      1399 AA; 166192 MW; 4CB3630FF708E6E7 CRC64;

```

```

Query Match      32.5%; Score 3093; DB 2; Length 1399;
Best Local Similarity 44.3%; Pred. No. 2.5e-151;
Matches 640; Conservative 284; Mismatches 421; Indels 100; Gaps 9;

```

```

Qy      469 MKKPATPRKVRDTSIQPSVKLYDSELYMQTCINPDPFAATTTIDPFLASTMYLDEQAVDR 528
      ::| | : | : | |||: ::| ||||| :|| |||||:| ::
Db      1 LKRTAVPCSLPPKSEEKRVFLYDSRHLKTLINPDPFAATTTCPFLTVMYLDERAFEQ 60
Qy      529 HQADFKKWLNALVSI PADLDADLNNKIDVGKLFNEVRNKELVVAPTKEEQSMNYLTKYRL 588
      :: |||||:||||| : | :|||:|::||| :||| | | ||
Db      61 YERQMKKWLNALVTIPADLDTEPNKPLDVGKLFDEVKSKELTAPTKELISSKYY-KTRL 119

```

## us-09-914-698-1.rup

Qy 589 ETLRKA AVELFFSEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLLCFNPL 648  
Db 120 NHLSRAGIALYTSEEIAMPLRKVAAQIEKQLLSLRTDRNLHLDLVLQRSILELLLCFNPL 179

Qy 649 WLRLGLEVVFGKEIQMQSNRDIVGLSTFILNRLFRNKCEEQRYSKAYTLTEEYAETIKKH 708  
Db 180 WLRLGLEVVFGQIELQSNRDIVGLSTFIIHRLFRDRYLEARNISKAYNLSRAYAEHMRKF 239

Qy 709 SLQKILFLLPFLDQAKQKRIVKHNPCLFVKKSPHKETKDILLRFSSELLANIGDITREL 768  
Db 240 TLRMVLFLLLFLDTAKRRKLIKHNPCLFVRNAPHKETKEILIRFASQLVSGIGDITKHM 299

Qy 769 RLGYVLQHRQTFLEDFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPASRLQR 828  
Db 300 RVGYVLSHKQSFLENYAFENLAVDLRDGVRLTRVMEIILLRDDLSASLRVPPISRLQK 359

Qy 829 IFNVKLALGALGEANFQLGGDIAAQDIVDGHREKTLSELLWQLIYKFRSPKFHAAATVLQK 888  
Db 360 IHNINLALVALEQADYKIAGNVTAKDIDGHREQTMSLLWQIVYKFRAPKFNAIAIVLQR 419

Qy 889 WWRRLHVLHVVIQRRIRHKELMRRHRAATVIQAVFRGHQMRKYVKLFKTERTQAAILQKF 948  
Db 420 WWRMNWLKVTISRRIEEKRALRREAAARTIQAAVRGYCVRVWYEAHRRQKLRAIVTIQRF 479

Qy 949 TRRYLAQKQLYQSYHSIITIQRWRAQQLGRQHRQRFVELREAAIFLQRIWRRRLF AKKL 1008  
Db 480 SRRYLAQKLAARRFSAIVRIQQWRTVRQMRQARERFLLCRKSAIVLQTSYRRYALGRKL 539

Qy 1009 LAAAE-----TARLQRSQKQQAAS 1029  
Db 540 LAAATLIGQIRAEAKHRHLQATIIQRSIKSYVIHRRHLHATVNGMVAFIRRKRLQNRSAK 599

Qy 1030 IQMQWRTYQLGRIQRHEFLRQDLIMFVQRRMRSKWSMLEQRKEFQQLKRAAINIQQRWR 1089  
Db 600 IQ-----AYQRMRIVRKEYLRSRSAACIQRRWRECMEARQLRNRFLLMRASAIRLQQQYR 655

Qy 1090 AKLSMRKCNADYLALRSSVLKVQAYRKATIQMRIDRNHYSLRKNVICLQQLRAIMKMR 1149  
Db 656 GWRQMRQDRHTYANARNLIVQVQRRWRGTLAMRKERANYRTLRRVTINVQRRFRARQAMQ 715

Qy 1150 EQRENYLRLRNASILVQKRYRMRQMIQDRNAYLRTKCIINVQRRWRATLQMRERKNY 1209  
Db 716 SEVERYRTLCKATVTLQQRFRANKAMMEQRQQYNSLRVATLCVQRRFRAQLSMRAARASY 775

Qy 1210 LHLQTTTKRIQIKFRAKREMKKQRAEFLQLKKVTLVVQKRRRALLQMRKERQEYLHLREV 1269  
Db 776 AKVRCAILTIQSQYRATLAMRHARDRFVTLRRCTITVQARFRAILAGRAAKQRYESIRKA 835

Qy 1270 TIKLQRRFHAQKSMRFMRKYRGTAQAVSCLQMHWRNHLRKRENSFLQLRQAATLQR 1329  
Db 836 TLHIQKWRATLEMRQVRSHYRRQCNAALTLQRSWGRVLLQKFRHDYLLYRGAATVLQR 895

Qy 1330 RYRARNMIKQLKSYAQLKQAATITRYRAKKAMQKQVVLVYQKQREAIKQRRYRGNL 1389  
Db 896 RYRALVQGRMVRREMQHCRWAAVTIQRRLRATLQMNDRKAFLQLRQSVLVVQRRFRAN- 954

Qy 1390 EMRKQIEVYQKQRAVIRLQKWWRSIRDMLCKAGYRRIRLSSLSIQRKWRATVQARRQR 1449  
Db 955 -----RACRVQVQYAAALKRSAITISHRWAA TLHMRQQR 988

Qy 1450 EIFLSTIRKVRMLMQAFIRATLLMRQQRREFEMKRAAVVIQRRFRARCAMLKARQDYQLI 1509  
Db 989 SDFLRKLSATVVMQRRYRAQRAKQAVQYERMRAAIVLLQKRYRAQRAMEKCRGRFLNL 1048

## us-09-914-698-1.rup

Qy	1510	QSSVILVQRKFRANRSMKQARQEFVQLRTIAVHLQQKFRGKRL----	MIEQRNCFQLLR	1565
Db	1049	KSASIVVQEFYRGRNMRHDRAAFIRLRESVLAIQRRFRGKLLTRQTVVDLR--	FEQIRR	1106
Qy	1566	SMPGFQARARGFMARKRFQALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRRQKHQGLLDI		1625
Db	1107	TVRGLQTYGRGVLARRAFLALLTPEYLERKRQKKAALRIQAWWRGAYHRKRYQTMQMRKI		1166
Qy	1626	RKRIAQLRQEAKAVNSVRCK--VQEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLLMWC		1683
Db	1167	AQQMVASRMAARRDPTIRLSNVSRCLRFLKTRFSSSEAIGILKRLERMSRLVPHLLMED		1226
Qy	1684	SEFMSTFCYGIMAQAIRSEVDKQLIERCSRIILNLARYNSTTVNTFQEGGLVTIAQMLLR		1743
Db	1227	AVFLSVFCYNMMAQAIRSEVDKILIEICARIILNLARFRGTKEQAFQEDGLVTVSQMLLR		1286
Qy	1744	WCDKDSEIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYMVRETKKLVARKEKMKQNAR		1803
Db	1287	WCDKDGCIGFSTLCTLLWVLAHDNKKKNAIRRYMISKDAIYMLRETKKLVQRKEKMRKNVQ		1346
Qy	1804	KP-----PPMTSGRYKSQKINFTPCSLPSLEPDFGIIRYSPYTFISSVYAFDTILCK		1855
Db	1347	RPVGCLVAPNPQLMR-----TVPSLEPDFGVNRSPYVVFYSSVFGFERVLQM		1393
Qy	1856	LQIDM	1860	
Db	1394	LEVDL	1398	